Best Practices in Data Analysis and Sharing in Neuroimaging using MRI

Thomas E. Nichols[1,*], Samir Das[2,15], Simon B. Eickhoff[3], Alan C. Evans[2,15], Tristan Glatard[2,14], Michael Hanke[4,16], Nikolaus Kriegeskorte[5], Michael P. Milham[6,17], Russell A. Poldrack[7], Jean-Baptiste Poline[8], Erika Proal[9], Bertrand Thirion[10], David C. Van Essen[11], Tonya White[12], B.T. Thomas Yeo[13]

[1]University of Warwick, Coventry, UK; [2]McGill University, Montreal, Canada; [3]Heinrich-Heine University Düsseldorf, Düsseldorf, Germany; [4]Otto-von-Guericke-University, Magdeburg, Germany; [5]MRC Cognition and Brain Sciences Unit, Cambridge, UK; [6]Child Mind Institute, New York, USA; [7]Stanford University, Stanford, USA; [8]University of California, Berkeley, USA; [9]NEUROingenia Clinical and Research Center, Mexico City, Mexico; [10]Inria, Paris-Saclay University, Paris, France; [11]Washington University in St. Louis, St. Louis, USA; [12]Erasmus University Medical Center, Rotterdam, The Netherlands; [13]National University of Singapore, Singapore; [14]Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada; [15]Montreal Neurological Institute, Montreal, Canada; [16]Center for Behavioral Brain Sciences, Magdeburg, Germany; [17]Nathan S. Kline Institute for Psychiatric Research, New York, USA;

*Correspondence: t.e.nichols@warwick.ac.uk.

Words: 3250 (excluding abstract)
Figures: 0
Tables: 1
References: 24

Abstract

Given concerns about the reproducibility of scientific findings, neuroimaging must define best practices for data analysis, results reporting, and algorithm and data sharing to promote transparency, reliability and collaboration. We describe insights from developing a set of recommendations on behalf of the Organization for Human Brain Mapping, and identify barriers that impede these practices, including how the discipline must change to fully exploit the potential of the world's neuroimaging data.

[Start of body text]

The advancement of science requires continuous examination of the principles and practices by which the research community operates. In recent years, this ongoing evaluative process has flagged concerns about the reproducibility of published research. From the early claim by John

Ioannidis in 2005 that “most published research findings are false”[1] to the recent work by the Open Science Collaboration, which attempted to replicate 100 psychology studies and succeeded in only 39 cases[2], there is mounting evidence that scientific results are less reliable than widely assumed.

Efforts promoting open science principles across fields (e.g.[3]) as a means of fostering transparency and reproducibility are valuable, but we also need efforts focusing specifically on human neuroimaging. To address this need the Organisation for Human Brain Mapping (OHBM) created the Committee on Best Practices in Data Analysis and Sharing (COBIDAS[4], http://www.humanbrainmapping.org/cobidas). This group was charged with creating a report that would compile best practices for open science in neuroimaging and distill these principles into specific research practices. The report was developed in collaboration with the OHBM community, which provided feedback on a draft and ratification of the final version.

In this commentary, we review the challenging issues that arose in the formation of the report, and identify initial success and the key remaining shortcomings in current practice.

What is Reproducibility?

Open science comprises a number of different goals and principles. The COBIDAS was specifically concerned with ‘Open Data’ and ‘Open Methodology’, both of which are in service of ‘Open Reproducible Research.’ An immediate challenge was to obtain a working definition of reproducibility. We considered a hierarchy of reproducibility concepts ranging from measurement and analytical stability, to broader notions of generalisability (Table 1). A very narrow notion of generalizability would be test-retest reliability on the same scanner, same subject, within 30 minutes, while a more extended notion would be using different scanners on the same subject with re-imaging occurring within 7 days. Generalization over analyses corresponds to re-analysis of the same data using identical or similar tools. One variant of this is “computational reproducibility”[5], where independent researchers re-analyse the data and compare their results. We also considered versions of generalizability corresponding to traditional scientific notions of “replication”, such as whether a result is stable over different samples of subjects or populations of subjects. The most challenging, and arguably most important form of generalizability is whether a finding additionally holds under variation in the stimuli and experimental methods. Underlying all of these concerns about reproducibility is how theory-building requires reproducible empirical phenomena, and thus a theory will only be as accurate and generalizable as the data that are used to inspire and/or test it.

Regardless of the precise scope of generalization, operationalising any of these versions of reproducibility requires explicit definitions of the outcome of interest, which in itself is a challenge. Previous efforts have found generally good measures of test-retest reliability of MRI for both voxel-wise and region of interest measures (e.g. [6-8]), but this is the most narrow notion of reproducibility. A large scale project to measure the generalisability of MRI findings across studies, akin to the Open Science Collaboration’s efforts in Psychology[2], has not been undertaken in neuroimaging; however the one effort that set out to reproduce brain structurebehavior correlations found only 1 of 17 findings were replicated[9], though this work is limited by small replication sample sizes. More work is needed in this area to better quantify the generalisability of MRI findings.

In short, quantifying "reproducibility" requires precisely defining the scope of variation being considered, the exact outcome that is being measured, and a metric of the stability of that outcome. The COBIDAS did not set out to estimate reproducibility, but was motivated to identify practices that can maximise analytical stability and generalizability of individual studies.

[Table 1 about here]

Prescribing best practice

Neuroimaging is a broad field, encompassing a range of approaches across a growing number of modalities. We restricted the scope of the COBIDAS report to include the range of all human neuroimaging using Magnetic Resonance Imaging (MRI), though most of the principles discussed can be applied to other modalities. We established 7 domains of practice, from experimental design and acquisition, through results reporting and data sharing. We quickly realised that it is neither feasible nor desirable to prescribe exactly how any one type of experiment should be conducted. For example, when looking at task fMRI, the optimal experimental design to use will depend on whether one is just trying to detect the presence of an effect or rather estimate the shape of the hemodynamic response function.

The one "practice" that can be universally commended is the transparent and complete reporting of all facets of a study, allowing a critical reader to evaluate the work and fully understand its strengths and limitations. This also facilitates subsequent research efforts by other investigators, who can exactly follow (or carefully manipulate) each aspect of a study. This includes conveying the "researcher degrees of freedom", by reporting other analytical paths applied unsuccessfully on the present data before arriving at the published results. Although formidable, the reporting checklists provided in the COBIDAS MRI report reflects the breadth and depth of information needed to ensure another researcher could replicate the work.

To further facilitate reproducibility, the COBIDAS report includes specific recommendations for statistical modelling, where specific (and common) bad practices have been identified[10,11]. We have also made concrete recommendations for data sharing, where practice is still evolving.

From solicited community input, we were struck by the emphatic and diverse views on the types of data to share. Some strongly felt it was essential to share the rawest form of the data from the scanner (DICOM format), while others felt that preprocessed, ready-to-analyze data should be shared; still others emphasized the utility of sharing extensively processed data linked to published figures. We evaluated the pros and cons of each form of data sharing; for example, while sharing preprocessed data can minimize the effort needed for reanalysis and speed advances based on new uses of the data, it may preclude alternate preprocessing options that facilitate new findings (e.g., more sophisticated image registration schemes, or changing the degree of spatial smoothing used). In the end, we endorsed the sharing of data in as many forms as is feasible.

Are we ready for open science in neuroimaging?

Brain imaging research is complicated, not only at the level of the conducting a study, but also at the level of sharing its results and data. The importance of thorough reporting of results is uncontroversial, and practices are improving, and the sharing of data to facilitate replication is increasingly viewed as essential. However, data sharing poses new challenges. Here we consider a number of concerns that investigators have with data sharing that impede adoption of open practices.

First, some individual researchers may assert ownership of their data and thus may not feel compelled to share. Counter to this is the drive for publically funded research to produce widely accessible data that can be reused and integrated into further research. Researchers may feel that sharing of data will result in a loss of competitive advantage, with other researchers swooping in to publish their planned studies based on the same data. The actual risk of this will depend on the data and hypotheses, but it should be weighed against the opportunity of new collaborations resulting from the sharing. These concerns can be alleviated by delaying the sharing or using a data-sharing repository with an embargo period.

Another fear is that, upon sharing data, other researchers will discover errors in an analysis or previously undiscovered problems with the data. As scientists, we are supposed to be objective arbiters of evidence and theory, but we are not infallible and must be ready to accept criticism and revise our claims when errors are discovered. Even when no errors are found, a re-analyses may support conclusions inconsistent with the original study. For controversial topics, there may also be adversarial reanalyses. We see no better way to advance understanding on a contested finding than to have as many researchers as possible puzzling over the data at hand. However, we need to develop a culture of constructive criticism that recognizes that errors are an inevitable part of scientific progress and protects individual researchers from inappropriately harsh consequences when honest mistakes are discovered.

A very practical concern, especially for junior investigators, is what is perceived as an unjustifiable cost of data sharing. Current incentives do not justify spending large amounts of time preparing data for sharing, as institutional promotion panels or grant reviewers currently do not adequately reward such efforts. Counter to this is the greater potential impact of a work when it may be cited not just for its scientific findings, but also when its data is reused in other works. Data description papers can document and provide credit for high-quality data acquisition efforts for the open community. We assert that if data sharing and open science priorities in general are to take hold, academic institutions, journals, and granting agencies are crucial for improving the incentives for open practices and developing ways to give appropriate credit for efforts in data sharing.

Finally there is the very real worry of failing to comply with human ethics provisions for protecting subject privacy. It can be argued that, once file headers are scrubbed of personally identifiable information and structural images have facial features obscured, that the data are completely anonymised and thus freely sharable. However individual ethics boards have varying views on this and it is best to write ethics consent documents explicitly with data sharing in mind. This topic would greatly benefit from leadership from national research organisations to seek consensus and then establish exactly what comprises anonymized brain imaging data. In particular, ethics boards often only try to minimize the risk to subjects when we are also obliged to maximize the benefit of our research to science and society, so as to honor the contribution of our subjects.[12] The future value of shared data must be considered in ethical decision making.

While studies lacking shared data and having opaque methodological detail are typical, some authors have embraced the challenges of sharing data and analysis methodology. Some recent examples that are particularly thorough and elegant include Waskom et al.[13] and Whitaker et al.[14], that published a complete array of analysis scripts for generating all figures and results in the paper (https://github.com/mwaskom/Waskom_JNeurosci_2014 and https://github.com/KirstieJane/NSPN_WhitakerVertes_PNAS2016, respectively), and Pernet et al.[15] that likewise shared raw data and analysis scripts, as well as all results maps in electronic form. From an organisational perspective, some labs are simply making open science a policy. Most recently the Montreal Neurological Institute announced that their work would be open, with all results and data made freely available at the time of publication[16].

These few examples demonstrate that some researchers are embracing open science principles, but do the tools exist to make it practical on a widespread basis?

Existing tools for open neuroimaging

There is an emerging ecosystem of open science tools for neuroimaging research. Before any data is collected, there are tools to assist in creating human ethics documents that maximise the ease of later data sharing, and for everything from experimental paradigm presentation, preprocessing to statistical modelling, neuroimaging benefits from numerous, free and well-supported software tools (see Supplementary Table 1 for an incomplete list). This constellation of tools could be seen as fuel for limitless researcher degrees-of-freedom, and indeed there is a need for the community to identify a set of ‘reference pipelines’ for common analyses. However, since each tool makes particular assumptions about neuroanatomical and neurophysiological processes, it is not possible to recommend the optimal analyses for every possible type of data and analysis objective. Only with user experience and reproducibility comparisons, will the field be able to identify what are the preferred analytical approaches.

There is a particular embrace of data sharing in the resting-state fMRI community. Since resting-state analyses methods remain in flux, sharing of this data has particular value as it allows future improvements in methods to be assessed and benchmarked relative to previous analyses. For resting and task fMRI and structural MRI, there are a number of projects that have led the way in this area, including the sibling projects FCON1000 and INDI[17], and the

Alzheimer's Disease Neuroimaging Initiative (http://www.adni-info.org). These have become invaluable tools for methodologists to apply novel image processing algorithms, not to mention the primary scientific outputs from these projects.

One promising new standard is the Brain Imaging Data Structure (BIDS)[18], a simple system for organising MRI data after conversion to the NIFTI format. BIDS provides a common, consistent directory hierarchy and naming system for files, as well as supporting ‘side car’ files for key associated data (like stimulus timing information for task fMRI). With a fixed standard for representing data, this has supported the creation of a number of “BIDS Apps”, self-contained programs that can automatically process data arranged according to BIDS. Simple, widely used standards such as this have the potential to dramatically reduce the effort required to exchange and share data.

New tools are set to dramatically advance computational reproducibility. A challenge to even something as simple as re-running the same data with the same code is the ever-changing versions of software and libraries that software depends on. The last five years has seen the growth of virtual machines and containers to share not just data but a complete environment for processing data. A virtual machine (VM) is an emulator of a computer, including its hardware, operating system and file system. It can be shared as a single file and when run, an entire computer system comes into existence based on a snapshot of the libraries and software interdependencies of one particular system. From within this VM, data can be run through a complete processing pipeline; with the original data of a study this will reproduce the results exactly, while new data can also be imported to evaluate the unique aspects of a pipeline. A downside to VMs is their gross size, as they are as large as any operating system. Containers are miniature VMs, lacking the full operating system but providing the specialised software and libraries required to execute a given task. The BIDS Apps mentioned above rely on such containers, encapsulating software packages large and small that alleviate installation of a myriad of software dependencies.

Open science tools are gaining traction. For example, the CBRAIN web-based analysis service supports over 260 collaborators in 20 countries; the COINS service currently hosts data on over 40,000 subjects for 643 studies; the LONI Pipeline has an average of 100,000 daily jobs from 200 different analysis workflows; the Neurovault repository hosts 450 public collections; and the FCP/INDI is openly sharing over 15,000 resting fMRI and structural MRI datasets.

Continuous improvement of research practices

Despite a seeming wealth of tools, there remain specific areas in the field of neuroimaging that need to be embraced to increase reproducibility. Aside from the importance of carefully reporting the study design, methods, and results mentioned above, we also identified priorities including archiving of statistical results, software engineering for reproducibility, and optimizing projects for generalizability.

In genetics, the routine sharing of "summary data" (SNP-level statistical results) has facilitated meta-analyses and methodological developments. For example, LD-score regression is a tool that can estimate genetic correlation using just Z-score summary data, and has had dramatic impact in a short timespan due to the availability of such results[19]. In brain imaging, we have no tradition of sharing summary statistics (i.e. images of T- or Z-scores, or images of percent change effect and standard errors). As a result the quality of meta-analyses are currently limited by their reliance on reported tables of maximum location coordinates, for which there is a substantial loss of information relative to the original statistic images[20]. In the current age, the costs of sharing such images of summary statistics (~1MB compressed), either through generic or dedicated repositories (e.g., NeuroVault.org, or BALSA, http://balsa.wustl.edu), are relatively minimal. As such, COBIDAS recommends the deposition of unthresholded statistical images into archival resources for all studies. Widespread adoption of this practice will dramatically increase our capacity for more precise meta-analyses, and allow more critical assessment of study results through exploration of the complete 3D image.

One foundation of computational reproducibility is modern software engineering practice. Whether a small set of scripts or a comprehensive end-to-end pipeline, neuroimaging data analysis depends on coding. Modern software engineering includes practices like version control and unit testing. Version control ensures that revisions of the code are identifiable and archived, and ideally is based on an open platform that allows wide inspection and input; unit tests verify the correctness of individual facets of the code, and can be set to automatically run each time the code is updated. This is not to say that every group should hire a programmer, but rather that every researcher writing scripts or code should obtain proficiency with basic software engineering skills and practices[21] (see Software Carpentry for basics instruction for non-programmers, http://software-carpentry.org/). With routine research grounded in well-written, less fragile code, it will be much easier to establish analysis pipelines that can both be reused within a lab and facilitate computational reproducibility verified by others.

Study designs have traditionally been optimised to maximise statistical power to detect differences between groups. With a growing emphasis on prediction, whether (e.g.) identifying early risk for psychosis or progression of a neurodegenerative disease, studies should be optimised for building predictive models that will generalise to the population of interest in yet-unseen data. Large multi-site studies that capture wide variation in human populations, as well as site-specific technical idiosyncrasies, are essential to build classifiers with good performance on new data. Whether obtained with prospectively optimized homogeneous acquisition and preprocessing strategies (e.g. Human Connectome Project and its successors[22]) or via larger but more heterogeneous, aggregate multisite approaches (e.g., FCON1000/INDI; ADNI, PING, and the upcoming ABCD Study) that have optimized image processing strategies determined retrospectively[23], generalisability of predictive models will be a key design objective and performance indicator going forward.

Beyond the investigator

Many of the practices advocated here and in the full COBIDAS MRI report require individuals to change the way they conduct research. Almost every such change requires an investment of time and resources. While we argue these have implicit rewards (e.g. shared data will never be lost when the post doc moves on), the advance of open science will require leadership at the institutional level. To provide appropriate incentives, universities and research centers need to explicitly consider the value of sharing of data and code as an unique research output in promotion and review exercises. Journals should require that papers' statistic images are archived, and promote papers with shared data, provide full analytical detail, and ideally share code or even executable containers or VMs. Foundations and granting agencies need to make data sharing a priority, recognizing and funding the explicit costs of data management required to make this happen. And finally professional organisations like OHBM should prioritize efforts in education to make open science practices accessible to all.

With the coordinated efforts of individual researchers, academic institutions, journals, granting agencies, and professional organisations, we can accelerate the drive towards open science and maximise the reproducibility of neuroimaging findings going forward.

Acknowledgements.

TEN is supported by the Wellcome Trust (100309/Z/12/Z) and NIH (R01 NS075066-01A1, R01 EB015611-01). BTTY is supported by Singapore MOE Tier 2 (MOE2014-T2-2-016), NUS (DPRT/944/09/14, R185000271720), NMRC (CBRG14nov007) and the NUS YIA. SBE is supported by the Deutsche Forschungsgemeinschaft (DFG, EI 816/4-1, LA 3071/3-1; EI 816/6-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme "Supercomputing and Modeling for the Human Brain" and the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 604102). MH was supported by funds from the German federal state of Saxony-Anhalt and the European Regional Development Fund (ERDF), project: Center for Behavioral Brain Sciences, and CRCNS BMBF/NSF (01GQ1411/1429999). NK was supported by the UK Medical Research Council and a European Research Council Starting Grant (261352). The authors declare no competing financial interests. ACE, SD and TG are supported by the Irving Ludmer Family Foundation and the Ludmer Centre for Neuroinformatics and Mental Health. TW was supported by a ZonMw TOP grant (91211021). RAP is supported by the Laura and John Arnold Foundation. MPM is a Phyllis Green and Randolph Cowen Scholar and is supported in part by the NIH (U01 MH099059; R01 AG047596)

REFERENCES

1. Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi:10.1371/journal.pmed.0020124
2. Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*(6251), aac4716. doi:10.1126/science.aac4716
3. Journals unite for reproducibility. (2014). *Nature*, *515*(7525), 7. doi:10.1038/515007a
4. Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., Yeo, B. T. T. (2016). Best Practices in Data Analysis and Sharing in Neuroimaging using MRI. bioRxiv, 54262. http://doi.org/10.1101/054262
5. Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226–7. doi:10.1126/science.1213847
6. Bennett, C. M., & Miller, M. B. (2013). fMRI reliability: influences of task and experimental design. Cognitive, Affective & Behavioral Neuroscience, 13(4), 690–702. http://doi.org/10.3758/s13415-013-0195-1
7. Schnack, H. G., van Haren, N. E. M., Brouwer, R. M., van Baal, G. C. M., Picchioni, M., Weisbrod, M., Hulshoff Pol, H. E. (2010). Mapping reliability in multicenter MRI: voxel-based morphometry and cortical thickness. Human Brain Mapping, 31(12), 1967–82. http://doi.org/10.1002/hbm.20991
8. Noble, S., Scheinost, D., Finn, E. S., Shen, X., Papademetris, X., McEwen, S. C., Constable, R. T. (2016). Multisite Reliability of MR-Based Functional Connectivity. NeuroImage. http://doi.org/10.1016/j.neuroimage.2016.10.020
9. Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. Cortex, 1–19. http://doi.org/10.1016/j.cortex.2014.11.019
10. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience, 12(5), 535–40. doi:10.1038/nn.2303
11. Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. NeuroImage, 40(2), 409–14. http://doi.org/10.1016/j.neuroimage.2007.11.048
12. Brakewood, B., & Poldrack, R. A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. NeuroImage, 82, 671–676. http://doi.org/10.1016/j.neuroimage.2013.02.040
13. Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition. Journal of Neuroscience, 34(32), 10743–10755. JOUR. http://doi.org/10.1523/JNEUROSCI.5282-13.2014
14. Whitaker, K. J., Vértes, P. E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., Bullmore, E. T. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. Proceedings of the National Academy of Sciences, 113(32), 201601745. http://doi.org/10.1073/pnas.1601745113
15. Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., Belin, P. (2015). The human voice areas: Spatial organization and inter-

individual variability in temporal and extra-temporal cortices. NeuroImage, 119, 164–74. http://doi.org/10.1016/j.neuroimage.2015.06.050

16. Owens, B. (2016). Montreal institute going “open” to accelerate science. Science. http://doi.org/10.1126/science.aae0265
17. Mennes, M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2013). Making data sharing work: The FCP/INDI experience. NeuroImage, 82, 683–691. http://doi.org/10.1016/j.neuroimage.2012.10.064
18. Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Scientific Data, 3, 160044. http://doi.org/10.1038/sdata.2016.44
19. Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics, 47(3), 291–295. http://doi.org/10.1038/ng.3211
20. Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., & Nichols, T. E. (2009). Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. NeuroImage, 45(3), 810–823. http://doi.org/10.1016/j.neuroimage.2008.12.039
21. Goble, C. (2014). Better software, better research. IEEE Internet Computing, 18(5), 4–8. http://doi.org/10.1109/MIC.2014.88
22. Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., Van Essen, D. C. (2016). The Human Connectome Project’s neuroimaging approach. Nature Neuroscience, 19(9), 1175–87. http://doi.org/10.1038/nn.4361
23. Abraham, A., Milham, M., Martino, A. D., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2016). Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. Neuroimage. http://doi.org/10.1016/j.neuroimage.2016.10.045
24. ISO. (2006). Statistics - Vocabulary and symbols. Part 2: Applied statistics. ISO 3534–2 (Second.). Geneva: ISO.

Table 1. A partial taxonomy of reproducibility in neuroimaging. For each type of reproducibility (row), the variable (column) that is held constant (•, bullet) or allowed to vary (D=different) is indicated; minus (-) indicates not applicable. Variations in the participant studied can be described in terms of the population they belong to (e.g. different patient groups or people from different cultures), or whether the same sample or a distinct sample of individuals is used. The MRI scanner used can be the same or not, and if the same participant sample is considered, the very same data can be used or new data can be acquired on the same or different days (visits) to the scanner. Experimental variation has many forms including the particular experimental design, but here we only consider stimuli. The type of stimulus used (stimulus population) may change, for example in a working memory experiment, letter stimuli might be replaced with shape stimuli; a more subtle change would be to use a different sample of stimuli of the same type, e.g. different particular shapes. The analysis method may vary; for example, with structural MRI for prediction of patient disease status, a linear discriminant might be used instead of a nonlinear support vector machine. Analysis code more narrowly reflects the particular implementation of a given method. Personnel conducting the research is another important source of variation, whether this is the experimenter or data analyst. Finally, note that the International Standards Organisation (ISO) has precise definitions of reproducibility[24] as indicated in the first three rows, but these capture only the minimal levels of generalizability.

Levels of generalization	**Participants**		**MRI Acquisition**			**Experiment**		**Analysis**		**Personnel**	
Generalization over measurements	**Population**	**Sample**	**Scanner**	**Visit**	**Data**	**Stimulus Population**	**Stimulus Sample**	**Method**	**Code**	**Experi-menter**	**Data Analyst**
ISO Repeatability e.g. 30-minute intra-scanner reliability	•	•	•	•	D	•	•	•	•	•	•
ISO Intermediate Reproducibility e.g. 7-day intra-scanner reliability	•	•	•	D	D	•	•	•	•	•	•
ISO Reproducibility e.g. 7-day inter-scanner reliability	•	•	D	D	D	•	•	•	•	•	•
Generalization over analyses											
Analysis Replicability	•	•	•	•	•	•	•	•	•	•	•
Collegial Analysis Replicability	•	•	•	•	•	•	•	•	•	•	D
Peng[5] Reproducibility	•	•	•	•	•	•	•	•	D	D	D
Generalization over materials and methods											
Near Replicability (different subjects)	•	D	•	-	-	•	•	•	•	•	•
Intermediate Replicability (different labs)	•	D	D	-	-	•	•	•	•	D	D
Far Replicability (different experimental & analytical methods)	•	D	D	-	-	•	D	D	D	D	D
Hypothesis Generalisability (different subject populations & types of stimuli)	D	D	D	-	-	D	D	D	D	D	D

Supplementary Table 1. An incomplete but illustrative list of free and well-supported tools for open science tools for neuroimaging. This table highlights analysis tools that can be scripted, allowing replicable analyses, as well as pipeline environments that bind together different software for replicable analyses, across heterogeneous software tools. The items under Data Sharing focus on tools to facilitate sharing and repositories that accept data. As repositories can have varying cost structures depending on the scale of data to be shared, we did not attempt classify as “free” or not; likewise, repositories generally do not comprise software that need to be downloaded, and we likewise did not attempt to classify by open source nature of the project. Results sharing tools either facilitate sharing or serve as repositories for shared results data. The Reproducibility Tools are a loose collection of resources that facilitate research using open science methods.

Resource	Type	Short Description	Free	Open Source	Link
Open Brain Consent	Consent	Ethics template oriented for neuroimaging data sharing	x	x	http://open-brain-consent.readthedocs.io
OpenSesame	Paradigm software	Graphical experiment builder	x	x	http://osdoc.cogsci.nl
PsychoPy	Paradigm software	Psychophysics software in Python	x	x	http://www.psychopy.org
Psychtoolbox	Paradigm software	Psychophysics Toolbox	x	x	http://psychtoolbox.org/
aa	Pipeline	Automatic Analysis, Matlab-based workflow tool	x (Matlab)	x	http://automaticanalysis.org
C-BRAIN	Pipeline	Web-based software for computationally intensive analyses	x	x	http://cbrain.mcgill.ca
CCS	Pipeline	Connectome Computation System, a pipline primarily for resting data	x	x	http://github.com/zuoxinian/CCS
C-PAC	Pipeline	Configurable Pipeline for the Analysis of Connectomes	x	x	http://fcp-indi.github.io
DPARSF/DPABI	Pipeline	Data Processing & Analysis for Brain Imaging, inlcuding resting-state fMRI	x	x	http://rfmri.org/dpabi
DTIPrep	Pipeline	Pipeline for diffusion weighted / diffusion tensor image data	x	x	http://www.nitrc.org/projects/dtiprep/
HCP Pipeline	Pipeline	Human Connectome Project Pipeline	x	x	http://github.com/Washington-University/Pipelines
LONI Pipeline	Pipeline	Cross-platform workflow tool for neuroimaging, genomics, bioinformatics	NC		http://pipeline.loni.usc.edu
LORIS	Pipeline	Web-based data and project management software for neuroimaging	x	x	http://loris.ca
NIAK	Pipeline	Llibrary of modules and pipelines for fMRI processing in Matlab/Octave	x	x	http://www.nitrc.org/projects/niak
NiDB	Pipeline	Neuroimaging database software that includes pipeline tools	x	x	http://github.com/gbook/nidb
NiPype	Pipeline	Neuroimaging in Python Pipelines and Interfaces	x	x	http://nipy.org/nipype
PANDA	Pipeline	Pipeline for Analyzing braiN Diffusion imAges	x	x	http://www.nitrc.org/projects/panda
SimNIBS	Pipeline	Simulation of Non-invasive Brain Stimulation	x	x	http://simnibs.de
AFNI	Scriptable Analysis	Neuroimaging analysis software for functional MRI	x	x	http://afni.nimh.nih.gov/afni
CONN	Scriptable Analysis	Functional connectivity toolbox, Matlab-based pipeline tool	x (Matlab)	x	http://www.nitrc.org/projects/conn
Connectir	Scriptable Analysis	Analysis software for Connectome-Wide Association Studies, based in R	x	x	http://czarrar.github.io/connectir
DiPy	Scriptable Analysis	Diffusion analysis pipeline using Python	x	x	http://nipy.org/dipy
Freesurfer	Scriptable Analysis	Neuroimaging analysis software for MRI, empahsis on surface-based analysis	x	x	http://surfer.nmr.mgh.harvard.edu
FSL	Scriptable Analysis	Neuroimaging analysis software for MRI	NC	x	http://www.fmrib.ox.ac.uk/fsl
MindBoggle	Scriptable Analysis	Automated labeling and shape analysis of brain images	x	x	http://www.mindboggle.info
SPM	Scriptable Analysis	Neuroimaging analysis software based in Matlab, for MRI, M/EEG, PET.	x (Matlab)	x	http://www.fil.ion.ucl.ac.uk/spm
Voxel	Scriptable Analysis	Mass-Univariate Voxelwise Analysis of Medical Imaging Data, based in R	x	x	http://cran.r-project.org/web/packages/voxel
BIDS	Data Sharing	Standard for organising MRI data and associated supporting data			http://bids.neuroimaging.io
COINS	Data Sharing	Web-based data management and analysis tool			http://coins.mrn.org
FCP/INDI	Data Sharing	Repository for resting state fMRI data			http://fcon_1000.projects.nitrc.org
Figshare	Data Sharing	Generic data sharing repository			http://figshare.com
LONI IDA	Data Sharing	Image data archive, repository for primarily neuroimaging data			http://ida.loni.usc.edu
LORIS	Data Sharing	Database for longitudinal imaging studies			http://bigbrain.loris.ca
NDA	Data Sharing	NIMH Data Archive, repository for data from NIMH-funded studies			http://data-archive.nimh.nih.gov
NITRC-IR	Data Sharing	Image repository for neuroimaging data			http://www.nitrc.org/ir
OpenfMRI	Data Sharing	Repository for task fMRI data, inlcuding all image and task paradam data			https://openfmri.org
PCP	Data Sharing	Preprocessed connectome project - pipelines for resting state data			http://preprocessed-connectomes-project.org
XNAT-Central	Data Sharing	Repository for raw MRI data			http://central.xnat.org
BALSA	Results Sharing	Sharing of surface-based statistical results	x		http://balsa.wustl.edu
NeuroVault	Results Sharing	Sharing tool for statistical maps	x	x	http://neurovault.org
NIDM	Results Sharing	Standard for exporting statistical results independent of the analysis tool	x	x	http://nidm.nidash.org
Docker	Reproducibility tool	Containerisation tool	x	x	http://www.docker.com
GitHub	Reproducibility tool	Version and issue tracking for software projects	x	x	http://github.org

Resource	Type	Short Description	Free	Open Source	Link
NeuroDebian	Reproducibility tool	Archive of research software packages for use on workstations & VMs	x	x	http://neuro.debian.net